

Durham Research Online

Deposited in DRO:

26 July 2019

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Awoke Ayele, Tadesse and Worku, Alemayehu and Kebede, Yigzaw and Zuma, Khangelani and Kasim, Adetayo and Shkedy, Ziv (2019) 'Model-based prediction of CD4 cells counts in HIV-infected adults on antiretroviral therapy in Northwest Ethiopia : a flexible mixed effects approach.', PLOS ONE, 14 (7). e0218514.

Further information on publisher's website:

<https://doi.org/10.1371/journal.pone.0218514>

Publisher's copyright statement:

©2019 Awoke Ayele et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

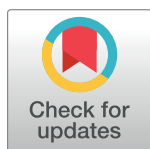
RESEARCH ARTICLE

Model-based prediction of CD4 cells counts in HIV-infected adults on antiretroviral therapy in Northwest Ethiopia: A flexible mixed effects approach

Tadesse Awoke Ayele^{1*}, Alemayehu Worku², Yigzaw Kebede¹, Khangelani Zuma³, Adetayo Kasim⁴, Ziv Shkedy⁵

1 Epidemiology and Biostatistics, University of Gondar, Gondar, Ethiopia, **2** School of Public Health, Addis Ababa University, Addis Ababa, Ethiopia, **3** Health Sciences Research Council, Pretoria, South Africa, **4** Wolfson Research Institute, Durham University, Durham, United Kingdom, **5** I-BioStat, Hasselt University, Diepenbeek, Belgium

* tawoke7@gmail.com



OPEN ACCESS

Citation: Awoke Ayele T, Worku A, Kebede Y, Zuma K, Kasim A, Shkedy Z (2019) Model-based prediction of CD4 cells counts in HIV-infected adults on antiretroviral therapy in Northwest Ethiopia: A flexible mixed effects approach. PLoS ONE 14(7): e0218514. <https://doi.org/10.1371/journal.pone.0218514>

Editor: Cristian Apetrei, University of Pittsburgh Centre for Vaccine Research, UNITED STATES

Received: February 11, 2018

Accepted: June 4, 2019

Published: July 10, 2019

Copyright: © 2019 Awoke Ayele et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors would like to thank University of Gondar for funding this research and TA received the funding. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Background

CD4 cell counts is widely used as a biomarker for treatment progression when studying the efficacy of drugs to treat HIV-infected patients. In the past, it had been also used in determining eligibility to initiate antiretroviral therapy. The main aim of this was to model the evolution of CD4 counts over time and use this model for an early prediction of subject-specific time to cross a pre-specified CD4 threshold.

Methods

Hospital based retrospective cohort study of HIV-infected patients was conducted from January 2009 to December 2014 at University of Gondar hospital, Northwest Ethiopia. Fractional polynomial random effect model is used to model the evolution of CD4 counts over time in response to treatment and to estimate the individual probability to be above a pre-selected CD4 threshold. Human subject research approval for this study was received from University of Gondar Research Ethics Committee and the medical director of the hospital.

Results

A total of 1347 patients were included in the analysis presented in this paper. The cohort contributed a total of 236.58 per 100 person-years of follow-up. Later the data were divided into two periods: the first is the estimation period in which the parameters of the model are estimated and the second is the prediction period. Based on the parameters from the estimation period, model based prediction for the time to cross a threshold was estimated. The correlations between observed and predicted values of CD4 levels in the estimation period were 0.977 and 0.982 for Neverapine and Efavirenz containing regimens, respectively;

while the correlation between the observed and predicted CD4 counts in the prediction period are 0.742 and 0.805 for NVP and EFV, respectively.

Conclusions

The model enables us to estimate a subject-specific expected time to cross a CD4 threshold and to estimate a subject-specific probability to have CD4 count above a pre-specified threshold at each time point. By predicting long-term outcomes of CD4 count of the patients one can advise patient about the potential ART benefits that accrue in the long-term.

Introduction

Human Immunodeficiency Virus (HIV) destroys CD4 cell count, which result to an increased plasma HIV RNA levels and experience an Acquired Immune Deficiency Syndrome (AIDS) in the long-run [1, 2]. Globally, a total of 36.7 million people are living with HIV [3] from whom two-thirds are from low and middle-income countries. Sub-Saharan Africa carries the highest burden of the diseases, 71% of the global total, but only about 12% of the world's population [4].

Despite the drawbacks in the development of a successful vaccine against HIV, the development of therapeutic regimens using drug combinations has significantly increased survival and reduced HIV-associated morbidities in HIV-infected individuals [5, 6]. Countries use World Health Organization (WHO) guideline to roll-out antiretroviral therapy (ART). CD4 cell counts was used to initiate standard first line regimens and to switch to second line during the previous time [7].

The primary goal of ART is to reduce HIV-related morbidity and mortality, prolong survival, improve the quality of life, restore and preserve immunologic function and prevent HIV-transmission [8]. The level of CD4 cell counts is routinely used for monitoring response to ART in HIV-infected patients. According to the CD4 cell count criteria, a patient would be eligible when his/her CD4 cells counts dropped below a given threshold value. The threshold value has been changed from less than 200 *cells/mm*³ in 2006 to less than 350 *cells/mm*³ in 2010 [9, 10]. The WHO 2013 guideline recommend that ART be initiated for all patients with CD4 count 500 *cells/mm*³ or less [11]. In 2015, WHO recommended HIV-treat all approach [12] based on two clinical trial outcomes [13, 14]. However, several studies are against earlier initiation of ART in patients who have high CD4 cell counts [15, 16]. This is because early exposure to ART may precipitate early evolution of resistance and unnecessary side-effects [17]. The standard therapy consists of three nucleoside reverse transcriptase inhibitors (NRTIs): Zidovudine (AZT), Tenofovir (TDF), and Stavudine (D4T) and one non-nucleoside reverse transcriptase inhibitor (NNRTI): nevirapine (NVP) or efavirenz (EFV) [18].

Limited studies were focused on modeling the CD4 cell counts trend over time for patients on ART especially in Sub-Saharan Africa. A study in Northwest Ethiopia [19] used semi-parametric mixed effect model to investigate CD4 counts response to treatments. Similarly, a study in eastern Ethiopia [20] used a linear mixed model which ignores the non-linear nature of the evolution. Another longitudinal study in Uganda [21] included cubic time effect to account the non-linear nature of CD4 count. In the current study, we proposed a flexible parametric modeling, the fractional polynomials framework [22–26], to predict a subject specific evolution of CD4 counts over time. We focus on two main issues: (1) an early prediction of CD4 counts under a specific treatment and (2) the estimation of the time to cross a given

CD4 threshold under treatment. The later allows us to compare the efficacy between different treatments.

This paper is structured as follows; Brief introduction is presented in the first part. In the second section, we describe the data and the proposed modeling approached is formulated. The proposed model is applied to the data and the results are reported in third section. A discussion is provided in the forth section.

Methods

Ethical clearance

A human subject research approval for this study was received from Institutional Review Board (IRB) of the University of Gondar. As the study was retrospective, the IRB waived that the research could be done based on record review without contacting the patients. Support letter was obtained from the medical director office of the hospital for retrieving retrospective data from the database and records. All the information was kept confidential, and no individual identifiers were collected.

Data

The data were collected from tertiary referral hospital in Northwest Ethiopia. The hospital serves a population of about 5 million people in the region and neighbours. The hospital has a voluntary counseling and testing clinic where both self-referred individuals and physician referred patients are tested for HIV. A total of 14, 000 HIV-infected individuals ever visited the ART clinic in the hospital, of whom 8927 ever started ART. Free ART service is offered to the population since 2005. The inclusion criteria for the study presented in this paper were (1): having at least two visits and one CD4 cell count measurements, (2) initiated ART within the study period, (3) initiated with either NVP or EFV as NNRTIs treatment groups, and (4) initiated with AZT NRTI backbones. In total, 2550 adults HIV/AIDS patients met the criteria, of whom 1347 initiated with AZT containing regimen. Available case analysis was used to handle missing data. The flow diagram of the data is presented in Fig 1.

Modeling CD4 cell counts using subject specific models

We considered a linear mixed effects model [27] given by

$$\mathbf{Y}_{it_{ij}} = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_{it_{ij}}. \quad (1)$$

Here, $\mathbf{Y}_{it_{ij}}$ ($i = 1, \dots, n_i, j = 1, \dots, m_i$) is n_i -dimensional response vector of log transformed CD4 counts of the i^{th} individual at time t_{ij} , \mathbf{X}_i is $n_i \times p$ and $n_i \times q$ dimensional fixed effect model matrices, whereas \mathbf{Z}_i is a random effects model matrices. The covariates Gender, WHO stage, age and NNRTI treatment were treated as fixed effect covariates and time (in month) is the random effect covariate. Note that t_{ij} is not the same for all patients which follow different schedule for their visits. The vector $\boldsymbol{\beta}$ is a p -dimensional vector of fixed effects and \mathbf{b}_i is a q -dimensional subject specific vector of random effects $\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D})$ and $\boldsymbol{\varepsilon}_{it_{ij}}$ is the random error term, $\boldsymbol{\varepsilon}_{it_{ij}} \sim N(0, \sigma^2 I_{n_i})$. In what follows we discuss the usage of the mixed effects model as a tool to obtain a model based prediction of the CD4 counts, and the estimation of the time to cross a CD4 count threshold under a given treatment regimen. According to individual and average evolution, CD4 cell count was non-linear over time. Thus, a flexible model formulation for the mean structure using fractional polynomial random effect model is presented.

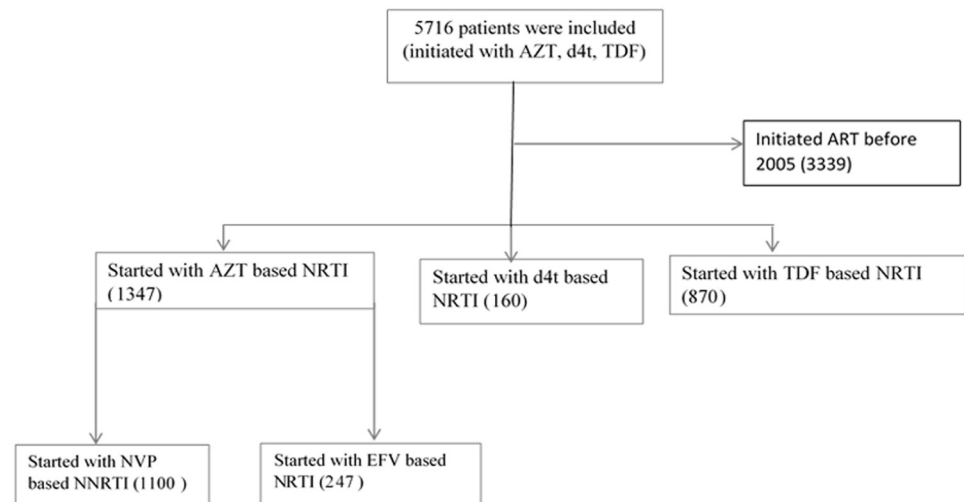


Fig 1. An illustration for the data flow diagram; A total of 5716 patients included, from which 58.4% initiated before 2005. A total of 1347 patients were included for final analysis (1100 NVP group and 247 EFV group).

<https://doi.org/10.1371/journal.pone.0218514.g001>

Model based prediction

The primary goal of the analysis presented in this paper is to obtain a model based subject specific prediction under a specific NNRTI treatment regimen of long term level of CD4 counts as early as possible. For this purpose, a two stage procedure was used. First, a mixed effect model is fitted using the data between 0 to 30 months. We term this period as *the estimation period*. In the second stage, the fitted model is used to predict the CD4 counts in the period of 31 to 68 months. We term this period as *the prediction period*. We determine the period for estimation and prediction by considering the length of the study period. Initially a period of 30 months of follow up was used as the period of prediction and analysis was conducted. Later sensitivity analysis was performed and the analysis supported the initial decision on the period of estimation.

The observed and predicted values in both estimation and prediction period are compared and their correlation was calculated in order to determine how good the model predicts the long-term CD4 counts. The procedure is illustrated in Fig 2a, where t_0 , t_p , and t_k represented the initiation time of ART treatment, estimation period and prediction period, respectively. Fig 3 illustrate profile for selected individuals.

Model based prediction of time to cross a pre-specified CD4 threshold

The linear mixed effect model formulated in Eq 1 can be used to predict the time that a subject will cross a pre-specified threshold level of CD4 counts. Let τ be a thresholds value, t_0 is the time to ART initiation, and T_τ the first time in which the subject CD4 counts crossed τ defined by

$$T_\tau = \min\{j \geq 1 : Y_{it_{ij}} \geq \tau\}, \quad (2)$$

Fig 2b illustrates schematically the trajectory of an individual who initiated ART at t_0 . In practice, three different threshold values of CD4 level were used for illustration: $\tau = \ln(200) = 5.2983$, $\tau = \ln(350) = 5.8579$, and $\tau = \ln(500) = 6.21461 \log(\text{cells}/\text{mm}^3)$. Natural logarithm transformation was used. The threshold value of $200 \text{ cells}/\text{mm}^3$ was recommended by WHO 2006 [9] as a criteria to initiate ART. In 2010, the criteria was modified to $350 \text{ cells}/\text{mm}^3$ [10]

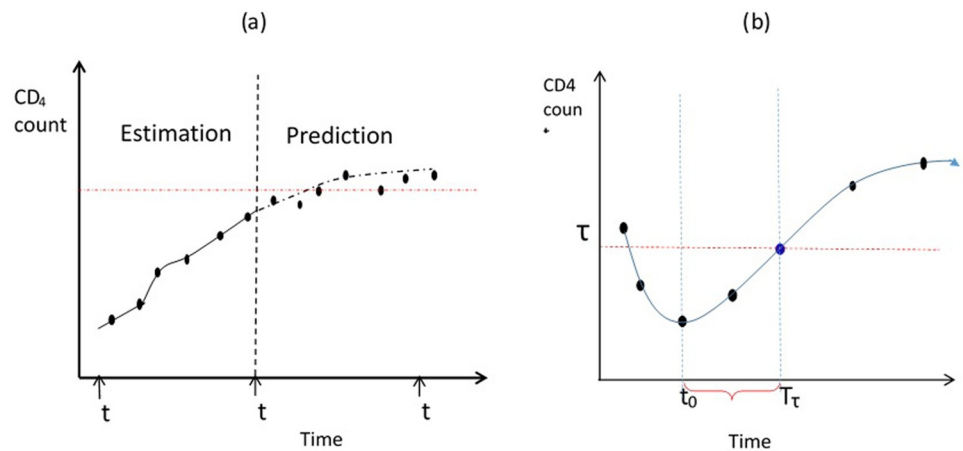


Fig 2. An illustration for model based prediction. Panel a: subject specific CD4 counts are divided into the estimation and prediction period. Panel b: an illustration of the time to cross a pre-specified threshold.

<https://doi.org/10.1371/journal.pone.0218514.g002>

and increased the threshold to 500 cells/mm³ in 2013 [11]. Since 2014, test and treat all approach is used irrespective of CD4 count level [12]. The current recommendation is viral load in order to monitor the progression of the virus and Ethiopia started implementing this recommendation in 2016 [28].

Based on the results presented in Reddy et al [29] the expected time for the i^{th} individual to reach a CD4 count greater or equal to the threshold τ can be expressed as

$$\begin{aligned} E(T_\tau) &= t_{i1}P(Y_{it_{i1}} \geq \tau) + t_{i2}P(Y_{it_{i1}} < \tau, Y_{it_{i2}} \geq \tau) \\ &+ t_{i3}P(Y_{it_{i1}} < \tau, Y_{it_{i2}} < \tau, Y_{it_{i3}} \geq \tau) + \dots \\ &= \sum_{j=1}^{\infty} t_{ij}S_{ij}. \end{aligned} \quad (3)$$

Here t_{ij} is the time corresponding to the j^{th} visit for the i^{th} individual, and S_{ij} is the probability of the i^{th} individual experiencing the event or stopping at t_{ij} . In practice, the infinite series will be truncated at a time point relevant to the study subjects. Condition on the random effects, the mixed model formulated in Eq 1 implies the so called conditional model [27], that is

$$Y_{ij}|\mathbf{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i, \Sigma_i).$$

Hence, the joint probability that form S_{ij} reduces to the product of the individual probabilities, which can be expressed as

$$\begin{aligned} S_{ij}(X_i, Z_i, \mathbf{b}_i, \boldsymbol{\beta}) &= P(Y_{it_{i1}} < \tau)P(Y_{it_{i2}} < \tau)P(Y_{it_{i3}} < \tau), \dots, P(Y_{it_{ij}} \geq \tau) \\ &= [\tilde{\phi}_{i1}(\tau)][\tilde{\phi}_{i2}(\tau)], \dots, [\tilde{\phi}_{ij-1}(\tau)][1 - \tilde{\phi}_{ij}(\tau)], \end{aligned} \quad (4)$$

where $\tilde{\phi}_{ij}(\tau)$ is a cumulative normal distribution with mean $X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i$ and variance σ^2 , that is

$$\tilde{\phi}_{ij}(\tau) = \phi\left(\frac{\tau - X_i\boldsymbol{\beta} - Z_i\mathbf{b}_i}{\sigma}\right), \quad (5)$$

Note that both fixed and random effects defined in the mixed model formulated in Eq 1 are used to calculate $\tilde{\phi}_{ij}(\tau)$. We elaborate this point in the next section.

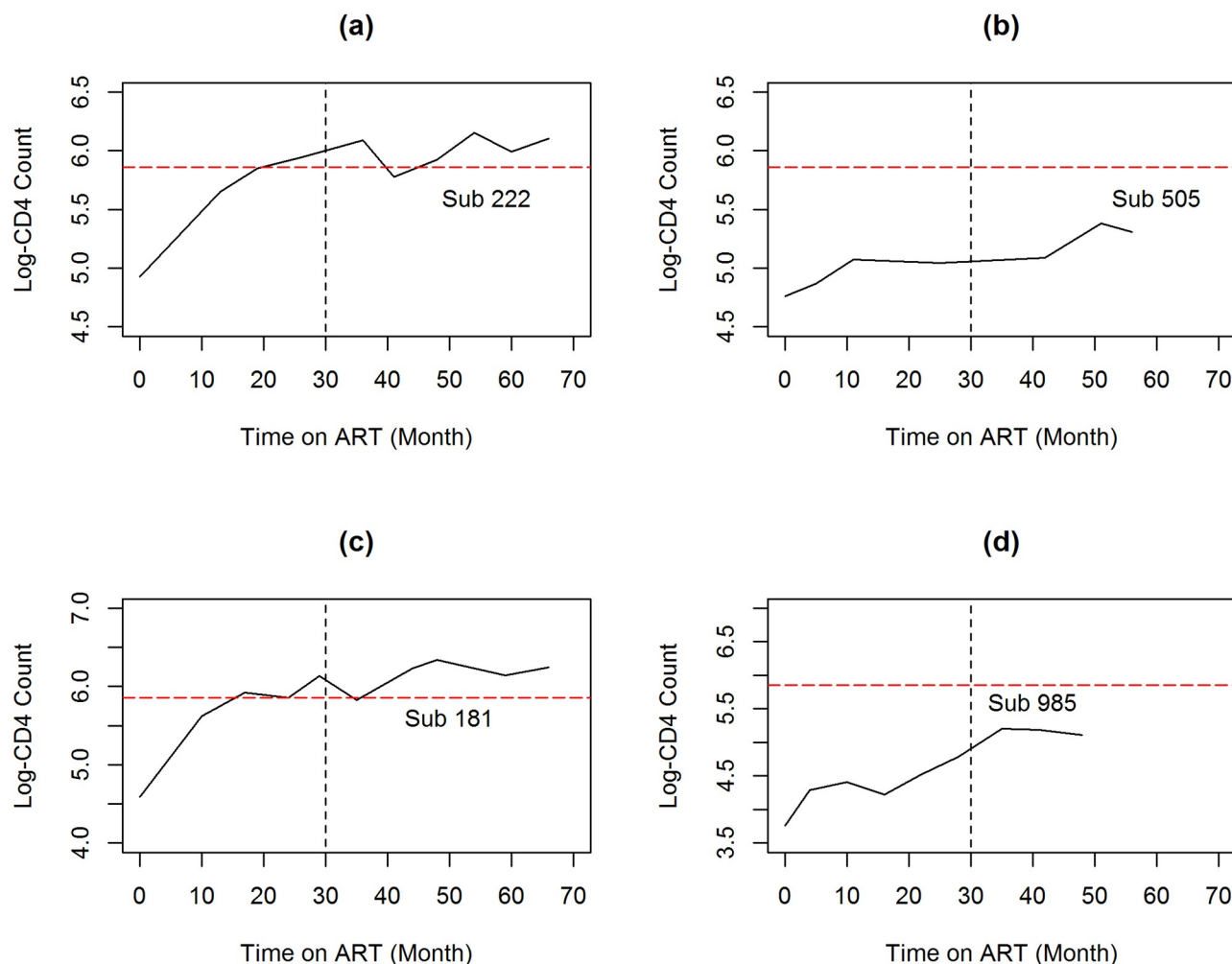


Fig 3. Individual profiles for patients who initiated ART with NVP (top panel) and EFV (bottom panel) containing regimens who cross and remain below the threshold.

<https://doi.org/10.1371/journal.pone.0218514.g003>

As shown by Reddy et al [29], the non-parametric bootstrap method can be used to compute standard errors and 95% confidence intervals for \hat{T}_τ . Four steps procedure is applied to compute the standard errors and confidence intervals:

1. Individual i is removed from the full dataset resulting $N - 1$ samples.
2. Sample $N - 1$ subjects with replacement from the dataset in step 1.
3. Append the data of individual i to the bootstrap sample.
4. compute \hat{T}_τ .

This procedure is repeated 1000 times.

Flexible modeling of the mean structure

The fractional polynomial model was proposed by Royston and Altman [22] as a flexible parametric approach to describe the dependency between a response of primary interest and continuous covariates. The responses of primary interest in the current application is the log

transformed CD4 cell counts and the covariate is time under ART treatment measured in months. The mean structure of an m order fractional polynomial model can be formulated as

$$\sum_{l=0}^m \beta_l H_l(t) + \sum_{l=0}^m b_{li} H_l(t), \quad (6)$$

where m is an integer, $p_1 \leq p_2 \leq \dots \leq p_m$ is a sequence of known powers and $H_l(t)$ is a transformation function given by

$$H_l(t) = \begin{cases} t^{p_l} & \text{if } p_l \neq p_{l-1}, \\ H_{l-1}(t) \times \log(t) & \text{if } p_l = p_{l-1}, \end{cases} \quad (7)$$

with $p_0 = 0$ and $H_0(t) = 1$. Note that there are two components in the mean structure. The first consists of the fixed parameters β_l and the later the subject specific parameters b_{li} .

For the analysis presented in this paper, both first ($m = 1$) and second ($m = 2$) order fractional polynomials mixed effect models were considered and compared. To select the power of the model, powers in the range $\{-2, -1.5, \dots, 2.5, 3\}$ were considered. Akaike Information Criteria (AIC, [30]) was used to choose the appropriate model. For a second order mixed effect fractional polynomial the mean structure is given by

$$f(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij}^{p_1} + (\beta_2 + b_{2i})t_{ij}^{p_2}, \quad (8)$$

Here, β_0 , β_1 , and β_2 are fixed effect parameters, and b_{0i} , b_{1i} , and b_{2i} are subject specific parameters. To compute cumulative probability above the threshold, the unknown quantities of Eq 5 were substituted by their estimate; $X_i \hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 t_{ij}^{p_1} + \hat{\beta}_2 t_{ij}^{p_2}$ and $Z_i \hat{b}_i = \hat{b}_{0i} + \hat{b}_{1i} t_{ij}^{p_1} + \hat{b}_{2i} t_{ij}^{p_2}$. R packages were used to perform all statistical analysis.

Results

Descriptive analysis

The median follow up period is 27.10 months (IQR = 12.10-43.10). The cohort contributed a total of 236.58 per 100 person years of follow up. The contributions were 237.11 per 100 person years and 234.28 per 100 person years by NVP and EFV containing regimens, respectively. The difference between the two treatment groups with regard to time contributed was 2.83 person years. The median number of repeated measurements was 3 (IQR = 2-6) with a maximum of 11 measurements per patient. At baseline, the absolute CD4 cell counts ranges between 2 and 2057 cells/mm³. To describe the pattern of CD4 change, different types of transformations are used [31]. In this study, log transformed CD4 cell count was used.

The evolution of CD4 cell counts over time for both treatment groups is shown in Fig 4 and reveal substantial variability between subjects. At baseline, 68.09% and 64.32% of the patients had CD4 cell counts below 200 cells/mm³ who initiated with EFV and NVP containing regimens, respectively. The percentage of patients who has CD4 cell count lower than 350 cells/mm³ increased to 94.53% and 91.97% for those who initiated with EFV and NVP containing regimen, respectively.

Model based prediction of CD4 cell counts: Estimation period 0-68 months

According to the AIC value, a second order fractional polynomial mixed effect model (FP2) was selected. The smallest value of AIC was obtained at ($p_1 = 0, p_2 = 0.5$) and ($p_1 = 0, p_2 = 0$) for NVP and EFV containing ART regimen, respectively (S1 and S2 Figs).

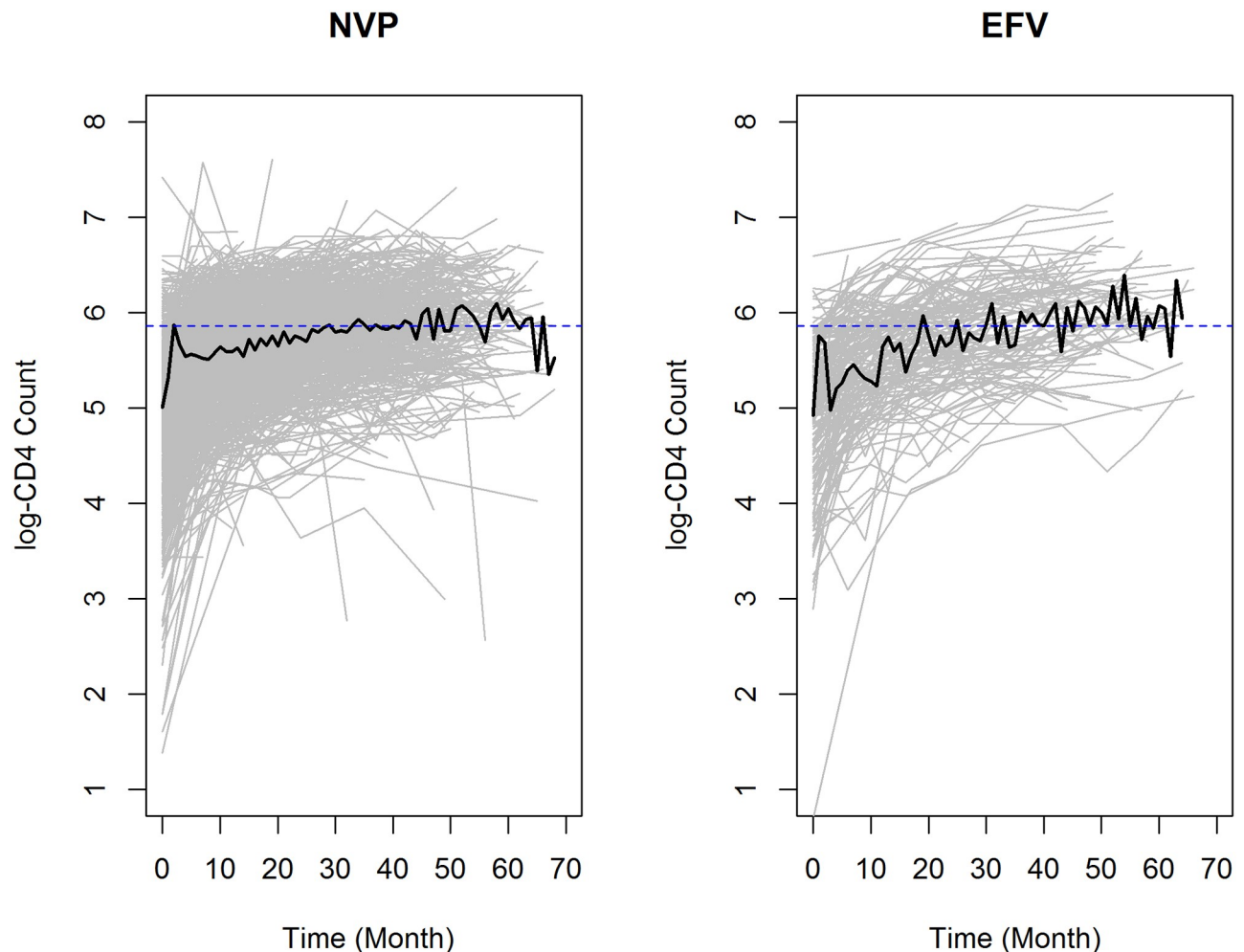


Fig 4. Individual and average profiles for patients who initiated ART with NVP (left panel) and EFV (right panel) containing regimens.

<https://doi.org/10.1371/journal.pone.0218514.g004>

Fig 5 shows the estimated mean profiles for the second order FP models for patients who initiated with NVP (left panel) and EFV (right panel) treatment regimen. The increase in log (CD4) cell counts from baseline was maintained until the end of follow-up period. Based on estimated FP models we can predict the log(CD4) counts for each subject using Eq (9), for subjects under NVP and EFV, respectively,

$$\begin{aligned}\hat{f}(t_{ij}) &= (5.22 + \hat{b}_{0i}) + (0.05 + \hat{b}_{1i})\log(t_{ij}) + (0.08 + \hat{b}_{2i})t_{ij}^{0.5}, \\ \hat{f}(t_{ij}) &= (5.05 + \hat{b}_{0i}) + (0.14 + \hat{b}_{1i})\log(t_{ij}) + (0.024 + \hat{b}_{2i})(\log(t_{ij}))^2,\end{aligned}\tag{9}$$

Here, \hat{b}_{0i} , \hat{b}_{1i} and \hat{b}_{2i} are the empirical Bayes estimates for the subject specific random effects. The fixed effect estimates are given in S1 Tables.

The density estimate for the distribution of the estimated and observed values for NVP and EFV at 6 and 12 months are shown in Fig 6 and indicates that the model is performing well in terms of Estimation at time points within the estimation period.

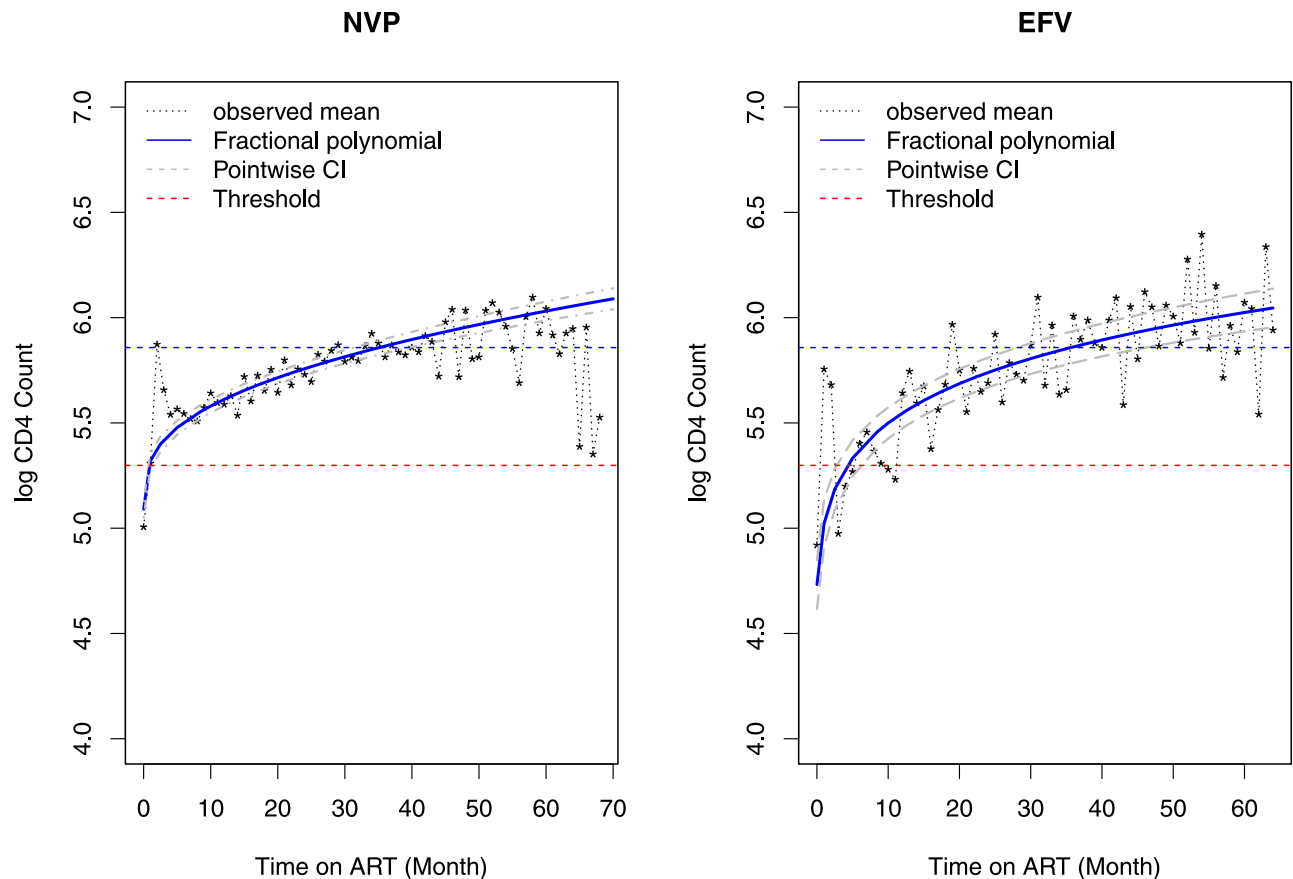


Fig 5. Model based predicted means for log(CD4) counts for NVP (left panel) and EFV (right panel) with observed average profile. Red dashed line: a threshold of 200. Blue dashed line: a threshold of 350.

<https://doi.org/10.1371/journal.pone.0218514.g005>

Observed and predicted values at all time points are presented in Fig 7 which reveals a strong positive correlation between the estimated and observed values (0.966 and 0.977 for NVP and EFV, respectively).

Model based prediction of CD4 cell counts: Estimation period 0-30 months

In the previous section model based prediction were obtained using FP2 models which were estimated using all data. In this section the data is divided into two periods. The first, 0-30 months, is used for the estimation of the model parameters while the second, 31 to 68 months, is used for prediction. Results obtained from a sensitivity analysis with different length for the estimation and the prediction period are given in the supplementary appendix of the manuscript (S1 Appendix). The analysis was performed for estimation periods of 0-24 months and 0-36 months. Fig 8a and 8b present the observed and predicted values of log(CD4) counts within the estimation period and reveal, similar to the previous section, high positive correlation (0.976 for NVP and 0.982 for EFV). Fig 8c and 8d display the predicted versus observed log(CD4) counts within the prediction period. Note that for this period the data were not used for the estimation of the model parameters. As expected, the correlations decrease to 0.805 and 0.742 for EFV and NVP, respectively.

The FP model estimated within the estimation period can be used to predict a subject specific last observed log(CD4) counts. This implies that for each subject, information about log

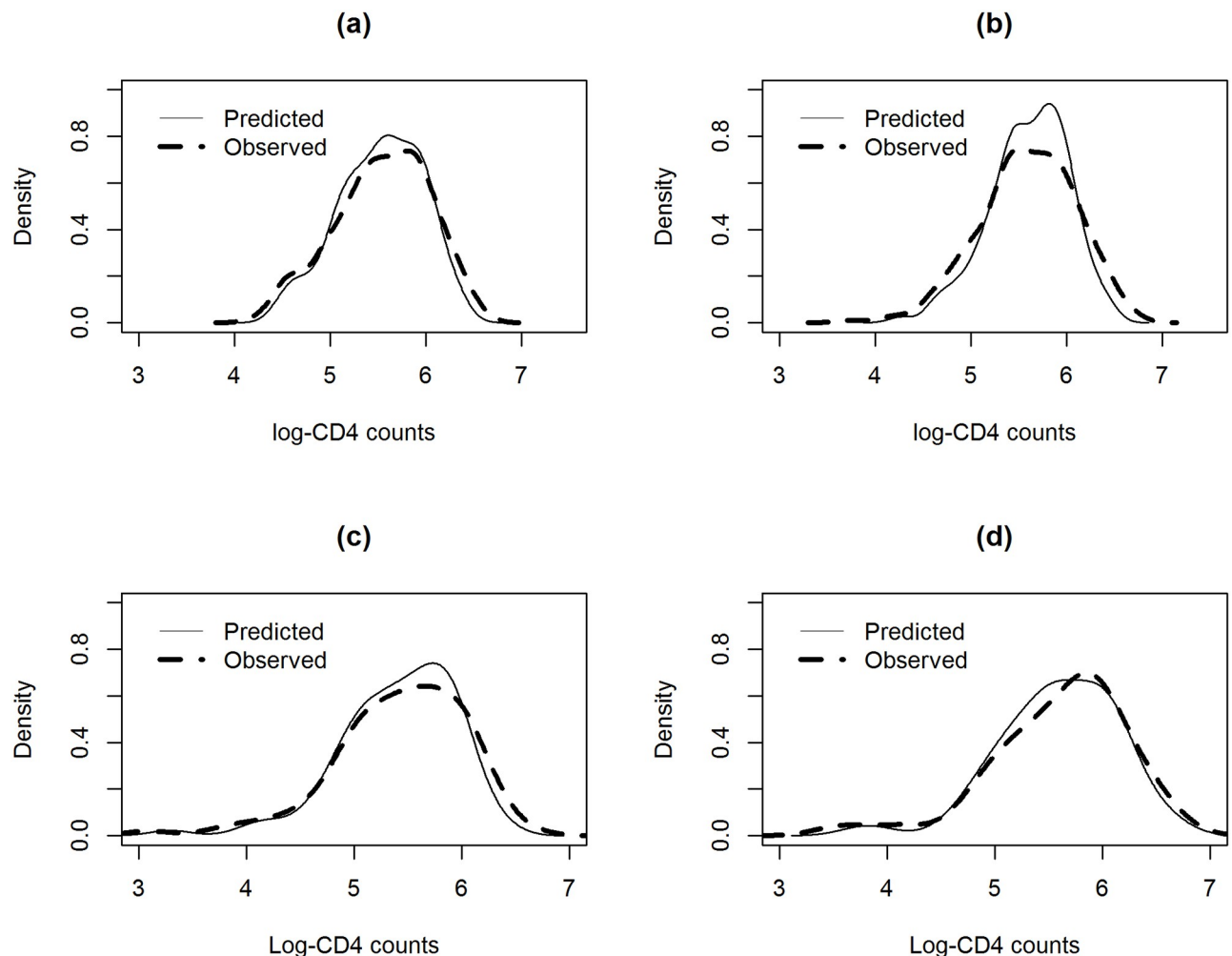


Fig 6. Density estimates for the distribution of the observed values (dashed line) and model-based estimation (solid line) Panel a: NVP, 6 months. Panel b: NVP, 12 months. Panel c: EFV, 6 months. Panel d: EFV, 12 months.

<https://doi.org/10.1371/journal.pone.0218514.g006>

(CD4) counts from the first 30 months of the treatment is used to predict the last observed log (CD4) count of the subject. [Fig 9a and 9b](#) present the observed and the predicted values. The correlations are equal to 0.764 and 0.808 for NVP and EFV, respectively.

Subject specific prediction of time to cross a pre specified CD4 threshold

The FP model was used to estimate a subject specific time to cross a pre specified CD4 threshold. Three different thresholds were considered (200, 350 and 500) to estimate a model-based time to cross the threshold which was considered as an event. Note that model based predictions are obtained using a model for which the parameters were estimated using data from the estimation period (i.e., 0-30 months).

[Fig 10](#) shows the observed and predicted log(CD4) counts for four selected patients. Panel *a* shows an example of a patient for which both the observed and predicted values cross the threshold within the estimation period. The time to cross the threshold is estimated to be 5.62 months. For patient presented in panel *b*, both the observed and predicted values cross the threshold within the prediction period (i.e 31-68 months). Panel *c* presents a patient for which

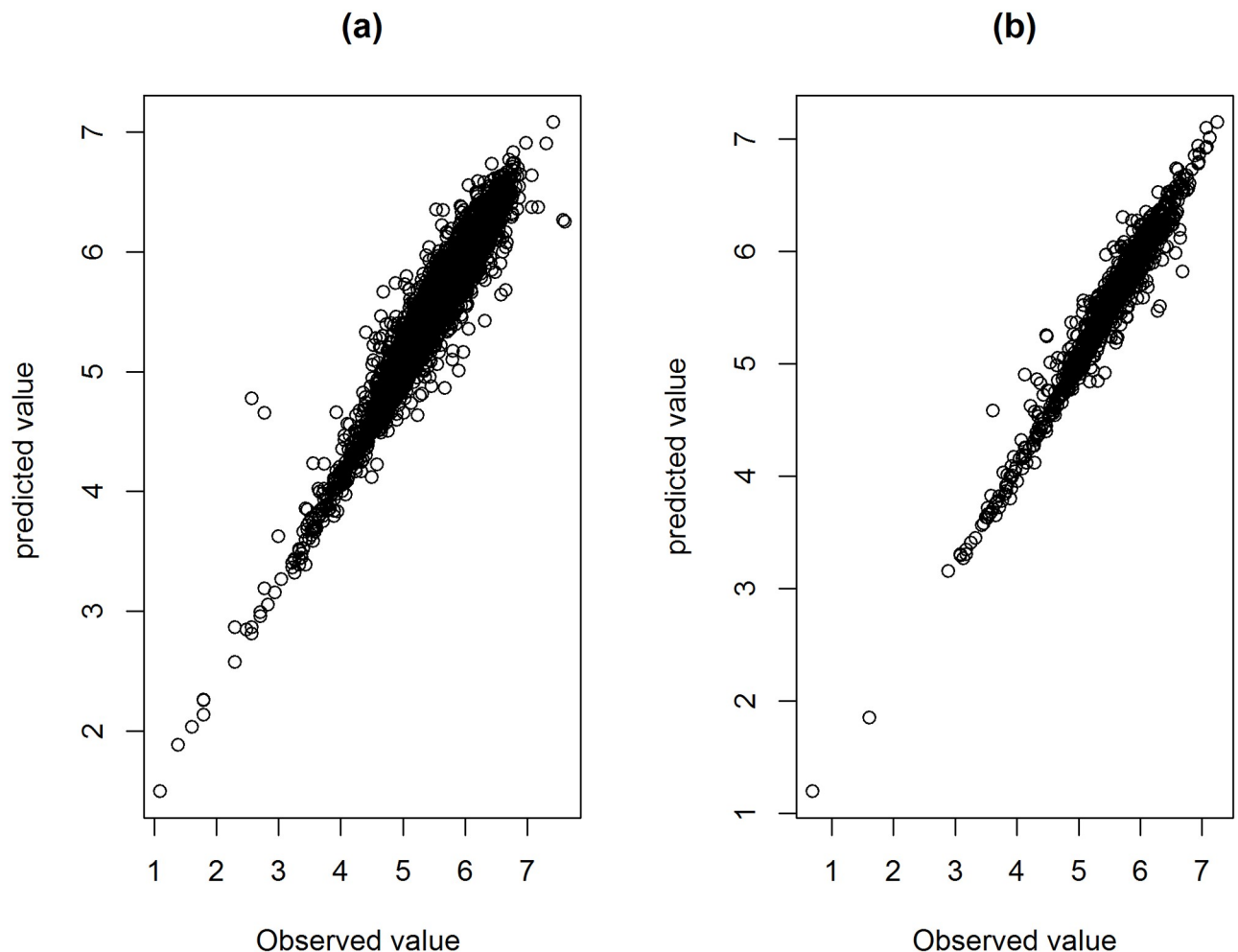


Fig 7. Observed versus model based predicted values for 0-68 months. Panel a: NVP. Panel b: EFV. Note that all data are used to estimate the predictive model.

<https://doi.org/10.1371/journal.pone.0218514.g007>

the observed log(CD4) count are below the threshold of 350 and the predicted time to cross the threshold is estimated to be 74.55 months. Panel *d* show an example of patients for which both the observed and predicted log(CD4) counts remain below the threshold of 350 until the end of the follow up period.

Fig 11 presents the kaplan-Meier curves for the time to cross the threshold of 200 and 350 cells/mm^3 . For the threshold of 200 $\text{CD4 cells}/\text{mm}^3$, there is a significant difference between EFV and NVP groups (p values for the log rank test is 0.0422). The median time to cross the threshold is estimated to be equal to 11.6 months (95% CI: 10.8-12.4) for NVP group and 15.0 months (95% CI: 12.7-17.3) for the EFV group. When the threshold increases to 350 (see Fig 11, panel *b*), the distribution of the time to cross the threshold of the groups were comparable (p-value = 0.52).

The predicted probabilities to cross a threshold of 350 cells/mm^3 at 60, 90 and 120 months were calculated according to Eq 3 and shown in Fig 12. For all time points, the probability to cross a threshold of 350 cells/mm^3 is higher for the EFV group. Although, the difference between the two treatment groups is higher for lower months, in the long run the two treatment groups had comparable result.

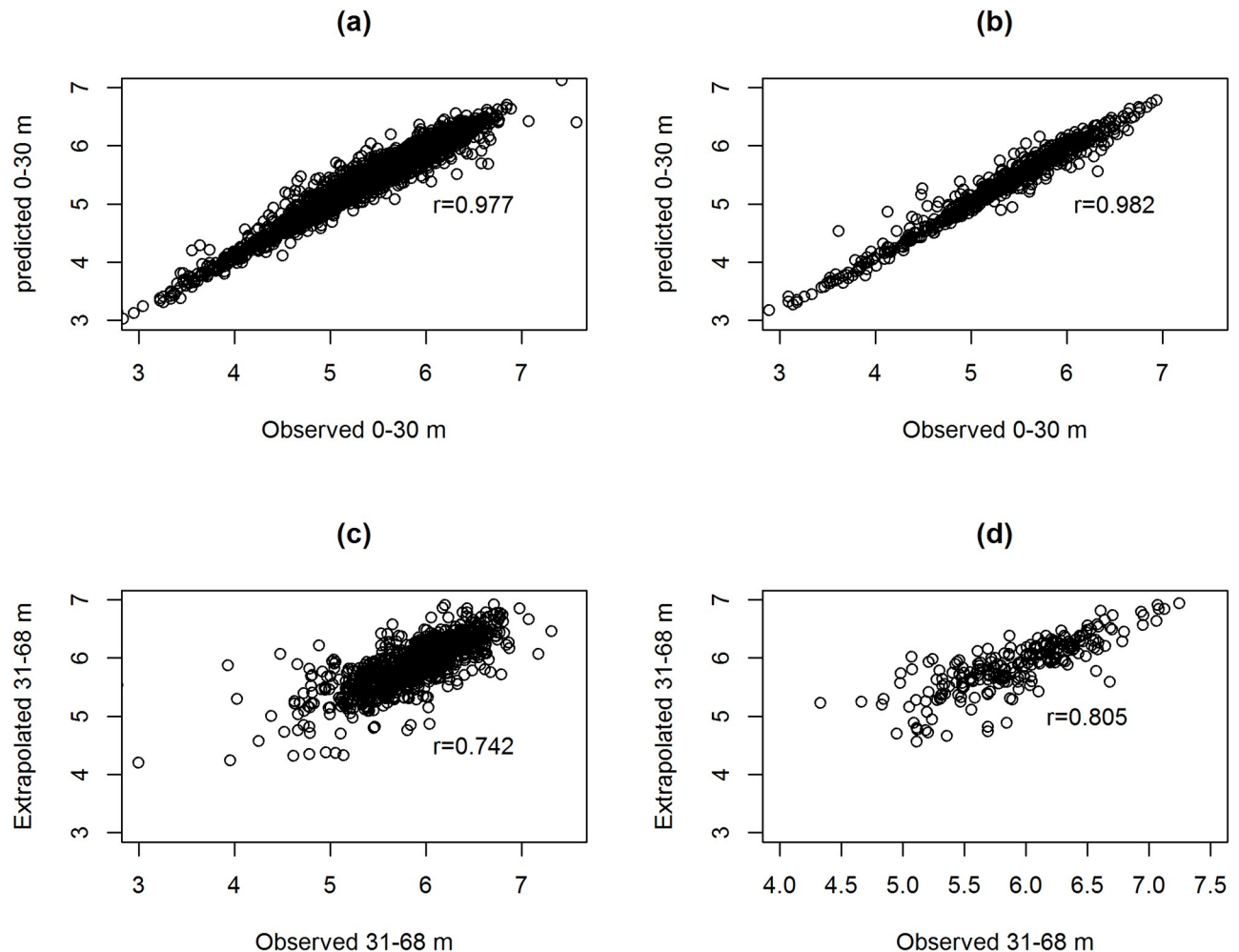


Fig 8. Observed versus model based predicted of CD4 counts obtained from a model that was estimated within the estimation period (0-30 months). Panel a and b: observed versus estimated values in the estimation period for NVP and EFV, respectively; Panel c and d are observed versus predicted values in the prediction period for NVP and EFV, respectively.

<https://doi.org/10.1371/journal.pone.0218514.g008>

For example, at 60 months, the probability to cross a threshold of 350 cells/mm^3 higher than 0.5 was calculated for 160 patients who were initiated on EFV (65% from the EFV sample) compare to 760 patients who were initiated with NVP (69% of the NVP sample). Fig 12d shows sorted predicted probabilities calculated using two models: (1) the black solid and green dash lines represent sorted probability for NVP and EFV regimens when the model is fitted for all available data, respectively and (2) the red and blue dash lines represent sorted probability for NVP and EFV regimens when the model is within the estimation period, respectively. We note that the two models lead to comparable probabilities in the estimation period and all observed dataset.

The procedures described in the methods section allows us to use Eq 3 to estimate both the time and the probability to be above a given threshold. For illustration we use 8 patients. Patients 2077, 1191, 67 and 783 are initiated with EFV containing regimen, while patients 44, 240, 747, and 252 are initiated with NVP regimen were selected. Table 1 shows the estimated time to cross the threshold and the corresponding 95% confidence interval for the selected patients.

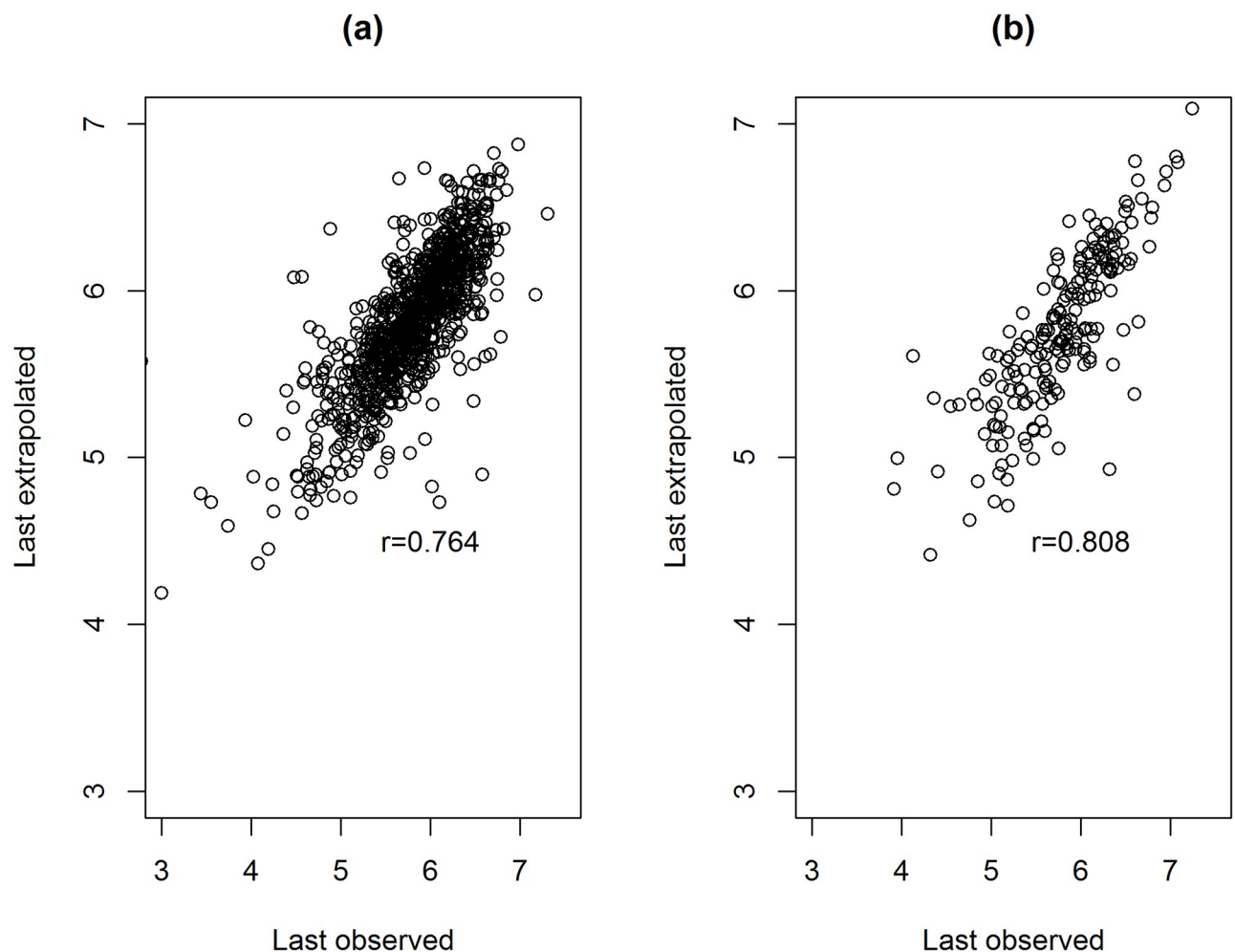


Fig 9. Last observed CD4 count and model based prediction based on a model which was estimated within the estimation period. Panel a: Patients who initiated with NVP containing regimen. Panel b: patients who initiated with EFV containing regimen.

<https://doi.org/10.1371/journal.pone.0218514.g009>

Patients 783 and 252 started ART when their CD4 counts dropped lower than 100 cells/mm^3 . The estimated time to cross 200 cells/mm^3 threshold for these patients were 20.282 (95% CI: 19.189-21.052) and 10.728 (95% CI: 10.473-11.027), respectively. Patients 67 and 747 were initiated ART at CD4 counts between 100 and 200 cells/mm^3 . These patients are expected to reach 200 cells/mm^3 threshold at 13.667 (95% CI: 12.737-14.495) and 6.711 (95% CI: 6.765, 7.096) months, respectively.

Patients 1191 and 240 have CD4 count greater than 200 cells/mm^3 when they start ART. The expected time to cross threshold 350 cells/mm^3 were 9.752 (95% CI: 9.456-10.081) and 12.102 (95% CI: 11.954-12.259), respectively. Patients 2077 and 44 were initiated ART at CD4 cell counts greater than 350 cells/mm^3 . The estimated times to reach the threshold 500 cells/mm^3 were 10.312 (95% CI: 9.897-10.890) and 16.717 (95% CI: 16.023-17.624), respectively. For each patient, the probability to cross the threshold of 350 cells/mm^3 , a different time points, was calculated according to Eq 5 and presented in Fig 13.

Discussion

In this paper we have shown that model based predictions are highly correlated with the observed values within the estimation period. Considerable relationship was observed in the

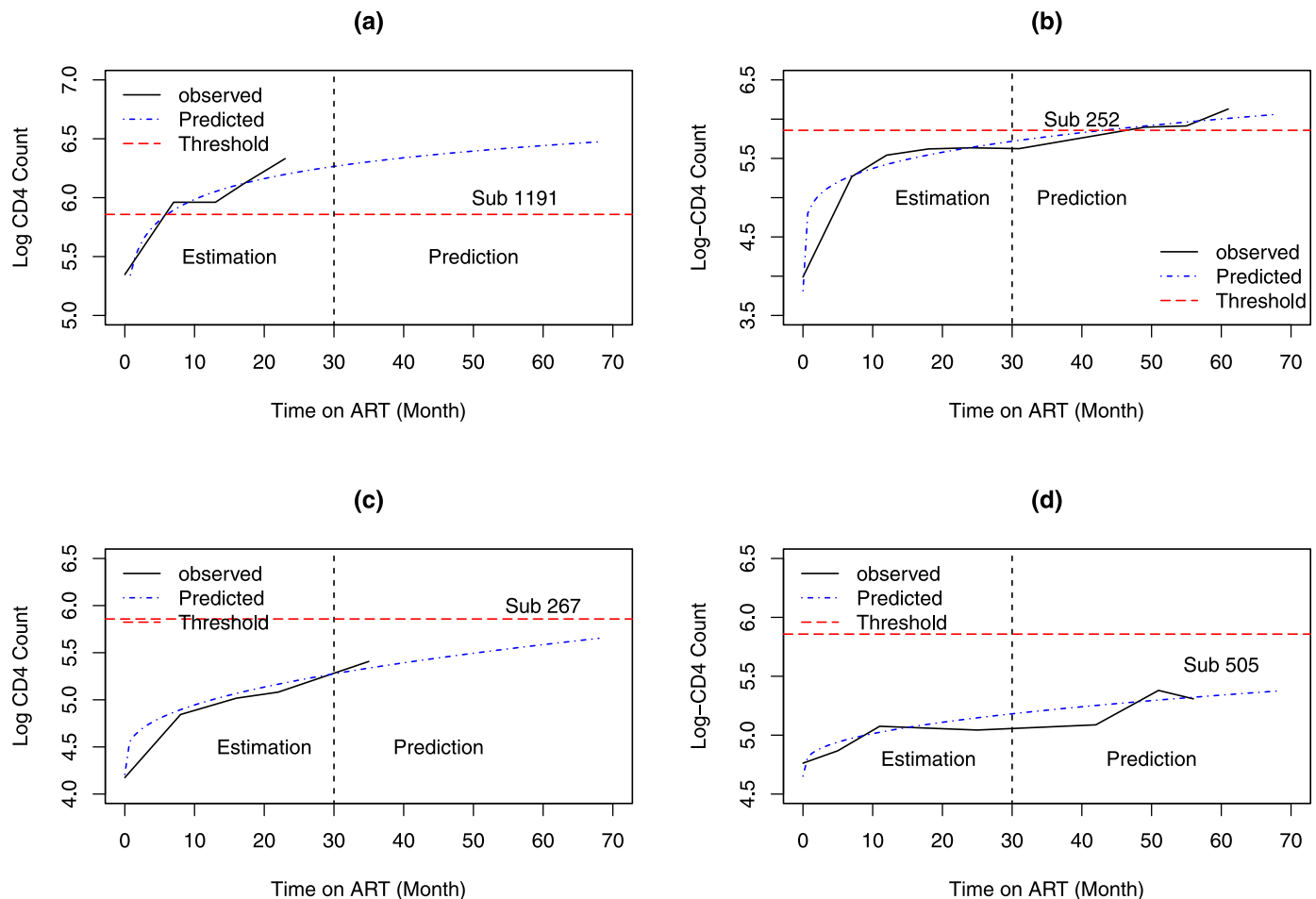


Fig 10. Observed and model based predicted CD4 counts for selected patients. Panel A: a patient for whom both observed and predicted values cross the threshold within the estimation period. Panel B: a patient for whom both observed and predicted values crossed the threshold after the estimation period. Panel C: a patient for whom the observed values are below the threshold and the predicted time to cross the threshold is estimated to be 74.55 months. Panel D: a patient for whom both observed and predicted values are below the threshold (350 cells/mm^3) during the study period.

<https://doi.org/10.1371/journal.pone.0218514.g010>

prediction period as well. This provides evidence that the proposed model can be used for long-term prediction of unobserved CD4 cell counts. The density estimates for the distribution of observed and predicted values supported these relations. Other studies used similar approach for long term prediction [32, 33]. The mixed effects FP model allows us to estimate the distribution of the time to cross a pre specified CD4 cell count threshold of interest and to use this distribution to compare between treatments. More than half (52.87%) of the patients who initiated ART at CD4 cell counts less than 200 cells/mm^3 cross the threshold in six months period after initiation. When the threshold is 350 cells/mm^3 , the proportion who crossed the threshold at 6, 24, 36 and 48 months were 13.88%, 40.14%, 52.79% and 61.65%, respectively. Several observational studies have reported that the probability of attaining elevated CD4 level can be sustained for at least 7 years and probably indefinitely [15, 34–36].

Patients who initiated at higher CD4 count get more pronounced CD4 cell count rise quickly than those who initiated at lower CD4 cell counts. Different studies showed that the baseline CD4 cell count influences the rate of immune reconstitution [34, 37–39]. These studies indicated that CD4 count at the time of ART initiation are critical determinants of the

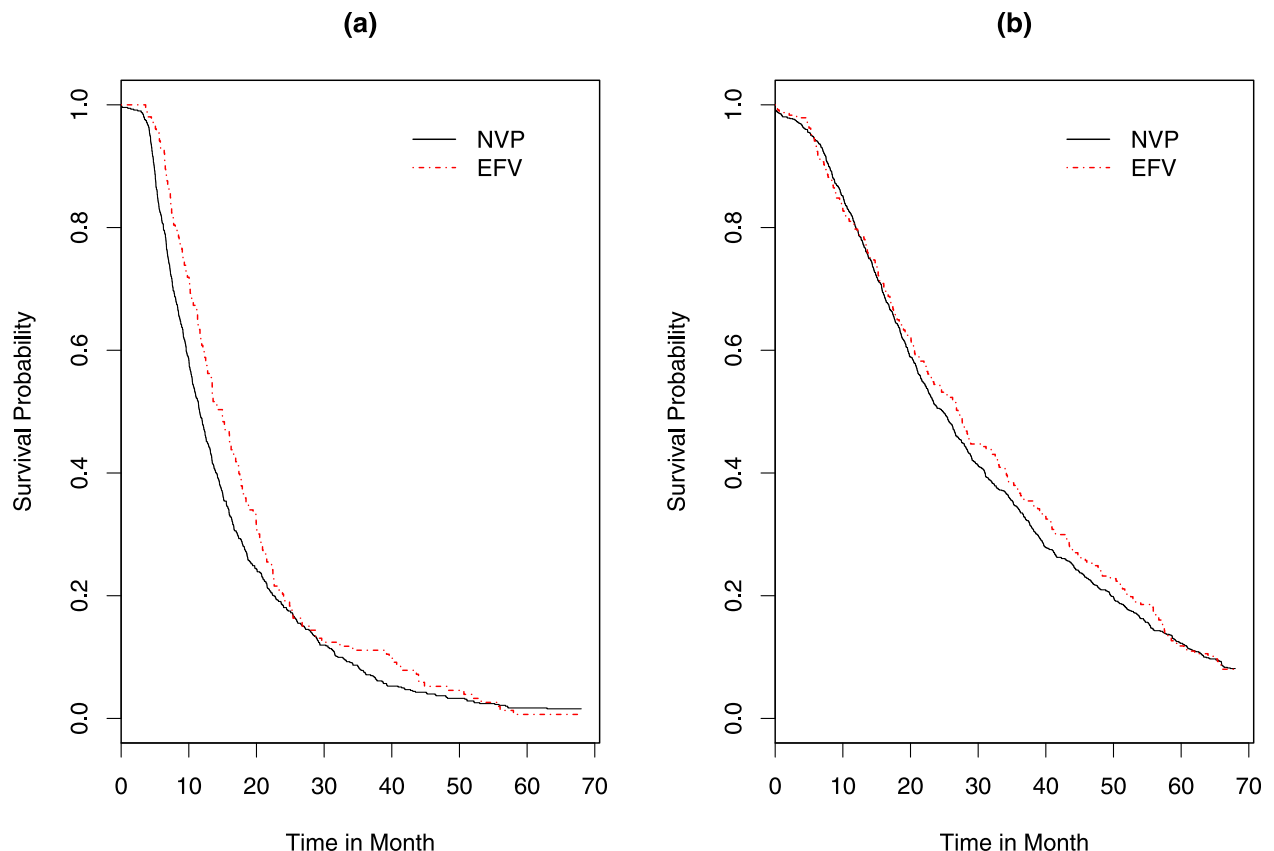


Fig 11. Kaplan-Meier curve for the time to cross a pre specified threshold of CD4 cells/mm³ by treatment group; Panel a: K-M curves for threshold of 200 cells/mm³ for EFV and NVP containing regimen. Panel b: K-M courses for a threshold of 350 cells/mm³ for EFV and NVP containing regimen.

<https://doi.org/10.1371/journal.pone.0218514.g011>

progression while under ART. This might be due to the fact that when the immune system is damaged, the risk of illness will increase.

We have shown that model based predictions of the time to cross a threshold reveal the similar patterns. There is difference in the time taken to cross the threshold among the two art regimens. The expected time to reach lower threshold is shorter for patients who initiated with NVP than EFV containing regimen. Whilst the expected time to reach higher threshold (350 cells/mm³) is shorter for EFV containing regimen which is also supported by other studies [40, 41]. This might be due to the high potency nature of EFV containing regimen.

The Kaplan-Meier survival curve also shows that the median time to cross the thresholds 200 CD4 cells/mm³ was shorter for patients who had been initiated with NVP as compared to EFV. Similar trend was reported by other studies [41, 42]. The possible reason is NVP has been used for patients with low CD4 level to reduce the side effect of EFV.

The limitation of this study was the unmeasured variables effect on the findings of the study which includes income, occupation, nutrition status and viral load. Note that the models presented in this manuscript included a limited number of covariates which were available in the database. Other covariates, if available in the database, such as TB status etc., can be included in the same way. The fractional polynomials model discussed in this paper was applied to a data based in which CD4 counts was the response variable for the assessment of patient progression under ART. In case that viral load measurements are available and used for progression assessment, the fractional polynomial modelling framework and the estimation

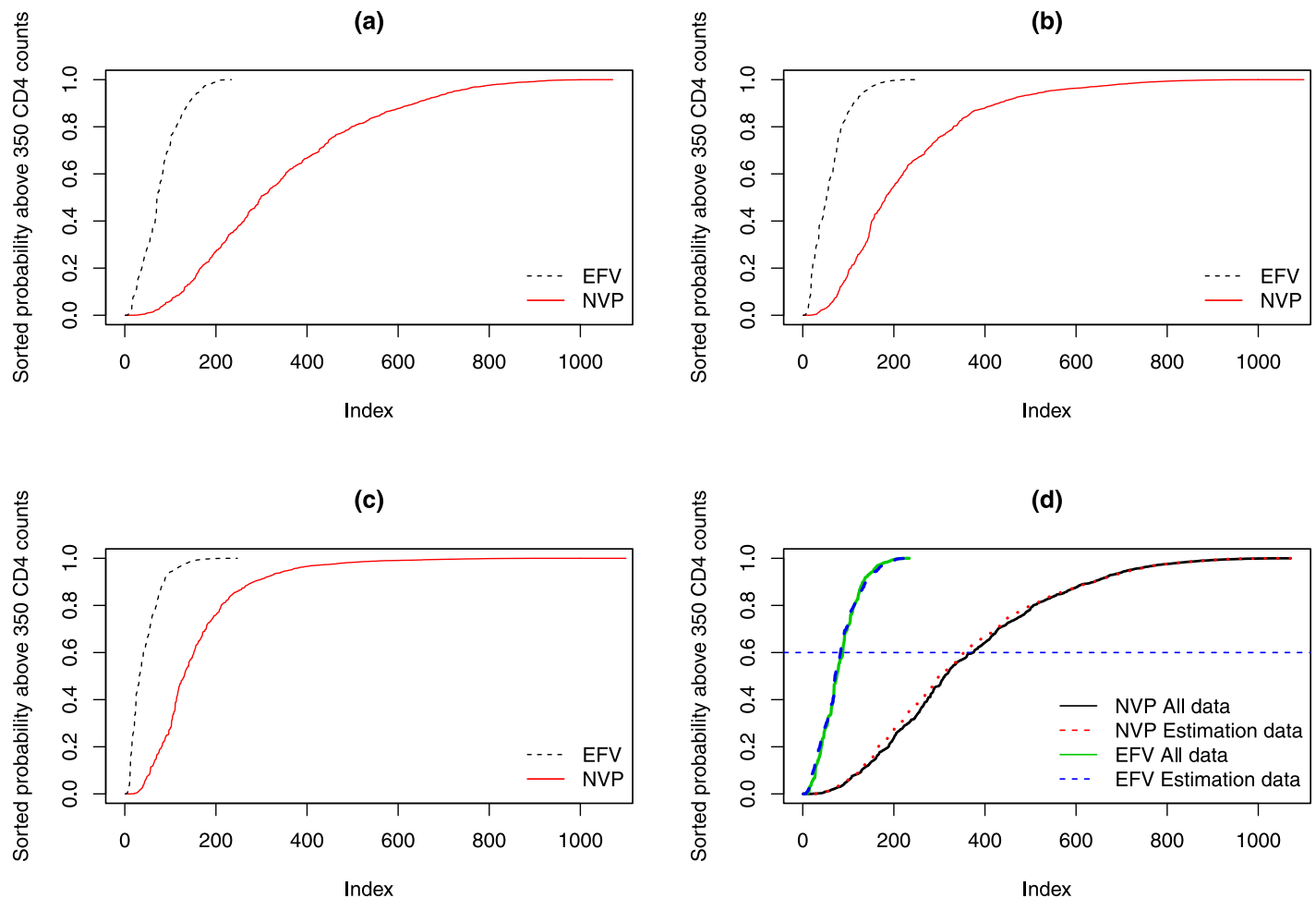


Fig 12. Sorted probabilities to cross a threshold of 350 CD4 counts. Panel a: sorted probability to cross the threshold at 60 months. Panel b: sorted probability to cross the threshold at 90 months. Panel c: sorted probability to cross the threshold at 120 months. Panel d: Sorted probability at 60 months obtained for models which were estimated within the estimation period and when using all data.

<https://doi.org/10.1371/journal.pone.0218514.g012>

Table 1. Estimated time to cross the threshold and 95% confidence interval for selected individuals.

PatientID	Baseline CD4	$\geq 200\text{cells/mm}^3$		$\geq 350\text{cells/mm}^3$		$\geq 500\text{cells/mm}^3$	
		$\hat{\tau}_{\tau}$	95%CI	$\hat{\tau}_{\tau}$	95%CI	$\hat{\tau}_{\tau}$	95%CI
EFV							
2077	425	2.8×10^{-3}	$(3.7 \times 10^{-4} - 1.151 \times 10^{-2})$	1.376	(1.045-1.763)	10.312	(9.897-10.890)
1191	210	2.789	(2.392-2.981)	9.752	(9.456-10.081)	19.786	(19.336-20.486)
67	126	13.667	(12.737-14.495)	42.918	(37.654-52.951)	79.928	(55.524-82.962)
783	67	20.282	(19.189-21.052)	51.361	(48.968-54.427)	82.1640	(71.872-85.256)
NVP							
44	490	2.5×10^{-3}	$(5.4 \times 10^{-4}, 9.0 \times 10^{-3})$	1.601	(1.177-2.142)	16.717	(16.023-17.624)
240	271	1.406	(0.975-1.501)	12.102	(11.954-12.259)	21.481	(21.001-22.085)
747	138	6.711	(6.765-7.096)	28.107	(27.259-29.968)	69.545	62.335-72.333
252	54	10.728	(10.473-11.027)	39.161	(38.496-40.721)	69.422	(67.492-71.566)

<https://doi.org/10.1371/journal.pone.0218514.t001>

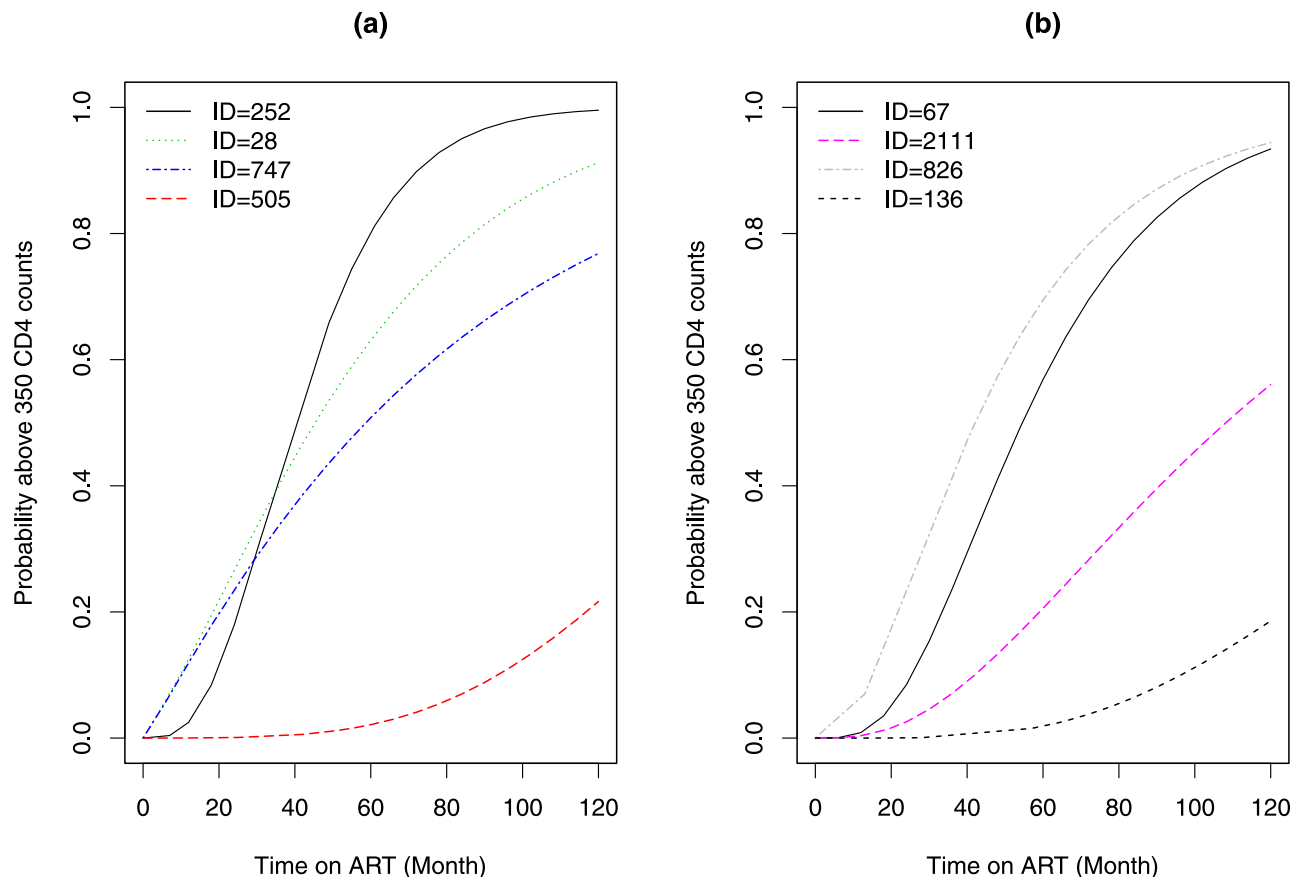


Fig 13. Probabilities to cross a threshold of 350 cells/mm³ for selected subjects. Panel a is for patients who initiated with NVP containing regimen, and panel b is for patients who initiated with EFV containing regimen.

<https://doi.org/10.1371/journal.pone.0218514.g013>

procedure discuss in this paper can be used to model the data. Moreover, additional covariates, if available, can be included in the model as well.

Conclusions

In conclusion, the model was used to estimate the probability of an individual to have CD4 count above a pre-specified threshold. By predicting the long-term outcomes of CD4 count of a patient one can advise patients about the potential ART benefits that accrue in the long term. Initiation of ART at higher CD4 cell counts has more benefit in achieving immunological success at a faster rate. Efavirenz containing regimen improves CD4 cells counts of the patient quicker than NVP containing regimen for higher baseline CD4 cell counts. Hence, those who have higher baseline CD4 cells count can be initiated with EFV containing regimen.

Supporting information

S1 Fig. Plots of powers for the first order fractional polynomial versus AIC value from the model. The left panel shows the plots of first degree powers against AIC for NVP and the right panel shows the plot of first degree powers against AIC for EFV.
(TIFF)

S2 Fig. Plots of powers for the second order fractional polynomial versus AIC value from the model. Panel (a) shows the plots of first and second powers against AIC for NVP and Panel (b) shows the plot of first and second powers against AIC for EFV.
(TIFF)

S1 Tables. Results of model comparison using akaike information criterion and model parameter estimates and their associated standard error using FP2.
(PDF)

S1 Appendix. Sensitivity analysis, when estimation period is 0-24 months and 0-36 months. Panel (a) and (b) show the plots of observed and predicted values in the estimation period for NVP and EFV regimens, respectively. Panel (C) and (d) show the plots of observed and predicted values in the prediction period for NVP and EFV regimens, respectively.
(PDF)

S2 Appendix. De-identified minimal data.
(CSV)

Acknowledgments

We would like to thank the University of Gondar and Hasselt University for the collaboration established to work on the project. The authors gratefully acknowledge the financial support from University of Gondar, Ethiopia. Our gratitude also goes to the staff of ART clinic of the hospital who collected the data and organized the database information according to the format.

Author Contributions

Conceptualization: Tadesse Awoke Ayele, Alemayehu Worku, Yigzaw Kebede, Khangelani Zuma, Adetayo Kasim, Ziv Shkedy.

Data curation: Tadesse Awoke Ayele, Ziv Shkedy.

Formal analysis: Tadesse Awoke Ayele, Ziv Shkedy.

Investigation: Tadesse Awoke Ayele, Yigzaw Kebede, Adetayo Kasim, Ziv Shkedy.

Methodology: Tadesse Awoke Ayele, Alemayehu Worku, Adetayo Kasim, Ziv Shkedy.

Project administration: Tadesse Awoke Ayele, Ziv Shkedy.

Software: Tadesse Awoke Ayele, Ziv Shkedy.

Supervision: Alemayehu Worku, Yigzaw Kebede, Khangelani Zuma, Adetayo Kasim, Ziv Shkedy.

Validation: Tadesse Awoke Ayele, Alemayehu Worku, Yigzaw Kebede, Khangelani Zuma, Adetayo Kasim, Ziv Shkedy.

Writing – original draft: Tadesse Awoke Ayele, Ziv Shkedy.

Writing – review & editing: Tadesse Awoke Ayele, Alemayehu Worku, Yigzaw Kebede, Khangelani Zuma, Adetayo Kasim, Ziv Shkedy.

References

1. Richard J, Horn T. The immune system, HIV and aging. New York: Treatment action group; 2014. Version 2.

2. Kestens L. The immune system and HIV: Virologic and immunologic basis of HIV infection + HIV pathogenesis; 2013.
3. Fact sheet—Latest statistics on the status of the AIDS epidemic;. Available from: <http://www.unaids.org/en/resources/fact-sheet>.
4. Kharsany AB, Karim QA. HIV infection and AIDS in Sub-Saharan Africa: current status, challenges and opportunities. *The open AIDS journal*. 2016; 10:34. <https://doi.org/10.2174/1874613601610010034> PMID: 27347270
5. Hoz JMDL, Bolaño L, Cárdenas O, González R, Sabbag J, Palacio L, et al. Characterization of treatment failure in HIV positive patients in the Colombian Caribbean region. *Colombia Médica*. 2014; 45(4):162–169.
6. Quinn TC. HIV epidemiology and the effects of antiviral therapy on long-term consequences. *AIDS*. 2008; 22(3):S7–12. <https://doi.org/10.1097/01.aids.0000327510.68503.e8> PMID: 18845925
7. Gilks CF, Crowley S, Ekpini R, Gove S, Perriens J, Souteyrand Y, et al. The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. *Lancet*. 2006; 368:505–10. [https://doi.org/10.1016/S0140-6736\(06\)69158-7](https://doi.org/10.1016/S0140-6736(06)69158-7) PMID: 16890837
8. Huldrych F, Judith A, Joseph J, Jennifer F, Amalio T, Constance A, et al. Antiretroviral Treatment of Adult HIV Infection 2014 Recommendations of the International Antiviral Society—USA Pane. *JAMA*. 2014; 312(4).
9. WHO. Antiretroviral drugs for treating pregnant women and preventing HIV infection in infants in resource limited settings; 2006 version.
10. WHO. Antiretroviral therapy for HIV infection in adults and adolescents. Recommendations for a public health approach; 2010 revision.
11. WHO. The use of antiretroviral drugs for treating and preventing HIV infection: Recommendation for a public health approach; 2013.
12. WHO. Guideline on when to start Antiretroviral therapy and pre-exposure prophylaxis for HIV; 2015.
13. Group TAS, et al. A trial of early antiretrovirals and isoniazid preventive therapy in Africa. *N Engl J Med*. 2015; 2015(373):808–822.
14. Group ISS, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N Engl J Med*. 2015; 2015(373):795–807.
15. Sabin C, Phillips A. Should HIV therapy be started at a CD4 cell count above 350 cells in asymptomatic HIV-1-infected patients. *Current Opinion in Infectious Diseases*. 2009; 22:191–197. <https://doi.org/10.1097/QCO.0b013e328326cd34> PMID: 19283914
16. Collaboration HC, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Annals of internal medicine*. 2011; 154(8):509. <https://doi.org/10.7326/0003-4819-154-8-201104190-00001>
17. Wood E, Hogg RS, Harrigan PR, Montaner PJS. When to initiate antiretroviral therapy in HIV-1-infected adults: a review for clinicians and patients. *Lancet Infect Dis*. 2005; 5(7):p407–414. [https://doi.org/10.1016/S1473-3099\(05\)70162-6](https://doi.org/10.1016/S1473-3099(05)70162-6)
18. Blas-García A, Apostolova N, Esplugues JV. Future Perspectives in NNRTI-Based Therapy: Bases for Understanding Their Toxicity;. Tang YW, editor. InTech; 2011.
19. Awoke T, Worku A, Kebede Y, Kasim A, Birlie B, Braekers R, et al. Modeling outcomes of first-line antiretroviral therapy and rate of CD4 counts change among a cohort of HIV/AIDS patients in Ethiopia: a retrospective cohort study. *PloS one*. 2016; 11(12):e0168323. <https://doi.org/10.1371/journal.pone.0168323> PMID: 27997931
20. Reda A, Biadgilign S, Deribew A, Gebre B, Deribe K. Predictors of Change in CD4 Lymphocyte Count and Weight among HIV Infected Patients on Anti-Retroviral Treatment in Ethiopia: A Retrospective Longitudinal Study. *PLOS ONE*. 2013; 8(4):e58595. <https://doi.org/10.1371/journal.pone.0058595> PMID: 23573191
21. Lubyayi L, Shkedy Z, Kajungu D, Mwehire D, Kirungi C, Kawuma E, et al. Evolution of CD4 Cell Counts over Time for HIV/AIDS Patients on Antiretroviral Therapy (ART) In Mildmay Uganda. *Biom Biostat Int J*. 2015; 2(2):00025.
22. Royston P, Sauerbrei W. Multivariate Model Building; A pragmatic Approach to Regression. Analysis Based on Fractional Polynomials for Modeling Continuous Variables;.
23. P R, DG A, W S. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. 2006; 25(1):127–141. <https://doi.org/10.1002/sim.2331>
24. Royston P, DG A. Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *ournal of the Royal Statistical Society*. 1994; 43(3):429–467.

25. FE H, KL L, DB M. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15(4):361–387. [https://doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4%3C361::AID-SIM168%3E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4)
26. Baneshi MR, Nakhaee F, Law M. On the Use of Fractional Polynomial Models to Assess Preventive Aspect of Variables: An Example in Prevention of Mortality Following HIV Infection. *International Journal of Preventive Medicine*. 2013; 4:414–419. PMID: [23671772](https://pubmed.ncbi.nlm.nih.gov/23671772/)
27. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. Springer Science & Business Media; 2009.
28. WHO. The use of antiretroviral drugs for treating and preventing HIV infection. Recommendation for public health approach; 2016.
29. Reddy T, Molenberghs Geert, Njagi EN, Aerts M. A novel approach to estimation of the time to biomarker threshold: applications to HIV. *Pharm Stat*. 2016; 15(6):541–549. <https://doi.org/10.1002/pst.1774> PMID: [27580636](https://pubmed.ncbi.nlm.nih.gov/27580636/)
30. Parzen E, Tanabe K, Kitagawa G. Selected Papers of Hirotugu Akaike. New York: Springer; 1998.
31. Taylor JM, Law N. Does the covariance structure matter in longitudinal modelling for the prediction of future CD4 counts? *Statistics in Medicine*. 1998; 17(20):2381–2394. [https://doi.org/10.1002/\(SICI\)1097-0258\(19981030\)17:20%3C2381::AID-SIM926%3E3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0258(19981030)17:20%3C2381::AID-SIM926%3E3.0.CO;2-S) PMID: [9819834](https://pubmed.ncbi.nlm.nih.gov/9819834/)
32. Michael L, Tilling K, Robertson N, Tremlett H, Zhuc F, Harding B, et al. A longitudinal model for disease progression was developed and applied to multiple sclerosis. *Journal of Clinical Epidemiology*. 2015; 68:1355–1365. <https://doi.org/10.1016/j.jclinepi.2015.05.003>
33. Aregay M, Shkedy Z, Molenberghs G, Davidc MP, Tibaldic F. Model-Based Estimates of Long-Term Persistence of Induced HPV Antibodies: A Flexible Subject-Specific Approach. *Journal of Biopharmaceutical Statistics*. 2013; 23(6):1228–1248. <https://doi.org/10.1080/10543406.2013.834917> PMID: [24138429](https://pubmed.ncbi.nlm.nih.gov/24138429/)
34. Ntokamunda J, Mukanyangezi MF, Uwizye CB. Effectiveness and Safety of Concurrent Use of First-Line Antiretroviral and Antituberculous Drugs in Rwanda. *Journal of Tropical Medicine*. 2014; p. 9.
35. Babiker A, Darbyshire J, Pezzotti P, Porter K, Prins M, Sabin C, et al. Short-term CD4 cell response after highly active antiretroviral therapy initiated at different times from seroconversion in 1,500 seroconverters. *J Acquir Immune Defic Syndr*. 2003; 32(3):303–10. <https://doi.org/10.1097/00126334-200303010-00010> PMID: [12626891](https://pubmed.ncbi.nlm.nih.gov/12626891/)
36. Minga A, Lewden C, Gabillard D, Bomisso G, d'Aquin Toni T, Emiène A, et al. CD4 eligibility thresholds: an analysis of the time to antiretroviral treatment in West African HIV-1 seroconverters. *AIDS*. 2011; 25(6):819–823. <https://doi.org/10.1097/QAD.0b013e32834625d3> PMID: [21412060](https://pubmed.ncbi.nlm.nih.gov/21412060/)
37. Smith CJ, Sabin CA, Youle MS, de Loes SK, Lampe FC, Cropley SMI, et al. Factors Influencing Increases in CD4 Cell Counts of HIV-Positive Persons Receiving Long-Term Highly Active Antiretroviral Therapy. *J Infect Dis*. 2004; 190(10):1860–1868. <https://doi.org/10.1086/425075> PMID: [15499544](https://pubmed.ncbi.nlm.nih.gov/15499544/)
38. Notermans D, Pakker N, Hamann D, Foudraïne N, Kauffmann R, Meenhorst P, et al. Immune reconstitution after 2 years of successful potent antiretroviral therapy in previously untreated human immunodeficiency virus type 1-infected adults. *J Infect Dis*. 1999; 180(4):1050–6. <https://doi.org/10.1086/315013> PMID: [10479130](https://pubmed.ncbi.nlm.nih.gov/10479130/)
39. Jacobson LP, Phair JP, Yamashita TE. Virologic and Immunologic Response to Highly Active Antiretroviral Therapy. *Current Infectious Disease Reports*. 2002; 4:88–96. <https://doi.org/10.1007/s11908-002-0072-5> PMID: [11853662](https://pubmed.ncbi.nlm.nih.gov/11853662/)
40. Forda N, Meintjes G, Vitoria M, Greened G, Chillerd T. The evolving role of CD4 cell counts in HIV care. *Curr Opin HIV AIDS*. 2017; 12:123–128. <https://doi.org/10.1097/COH.0000000000000348>
41. Teeranaipong P, Sirivichayaku S, Mekprasan S, Ohata PJ, Avihingsanon A, Ruxrungtham K, et al. Role of Rilpivirine and Etravirine in Efavirenz and Nevirapine-Based Regimens Failure in a Resource-Limited Country: A Cross-Sectional Study. *PLoS One*. 2016; 11(4):e0154221. <https://doi.org/10.1371/journal.pone.0154221> PMID: [27120449](https://pubmed.ncbi.nlm.nih.gov/27120449/)
42. van Letha F, Andrews S, Grinsztejn B, Wilkins E, Lazanase M, Langea J, et al. The effect of baseline CD4 cell count and HIV-1 viral load on the efficacy and safety of nevirapine or efavirenz-based first-line HAART. *AIDS*. 2005; 19:463–471. <https://doi.org/10.1097/01.aids.0000162334.12815.5b>